Predicting Neighborhoods to Open a New Shopping Mall in Bangalore, India

Predicting Neighborhoods for opening a new shopping mall is of importance to Property developers and land investors.

- The IT capital of India is facing rapid urbanization with an influx of people from all across the country.
- This necessitates greater infrastructure to cater to the diverse needs of the people and ensuring that one particular neighborhood doesn't suffer from oversupply while another suffers from undersupply is of paramount importance.
- So, property developers are always on the lookout for viable and most profitable land to build new malls to capitalize on customer needs.

Data Sources and Cleaning

- **Wikipedia** (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore)
- **Geocoder**
- Gives the geographical latitude and longitude coordinates of the neighborhoods using the *Geocoder* package.
- **Foursquare API**
- This gives the nearest the nearest venues within 4km of a neighborhood and in particular, the 'Shopping Mall' category.

Exploratory Data Analysis

```
# append the data into the list
for i in range(0,8):
    for row in soup.find_all("table", class_="wikitable sortable")[i].find_all("td"):
        neighborhoodList.append(row.text)
neighborhoodList = neighborhoodList[::3]
neighborhoodList = ([s.strip("\n") for s in neighborhoodList]) # remove \n from the string borders

neighborhoodList[0] = 'Catonment area, Bangalore'  # Clarifying to Geocoder so that it doesn't take the latitude and longitude c

neighborhoodList.append('Attibele')            # These are the additional areas in Bangalore Urban District that weren't properl
neighborhoodList.append('Chandrapura')         # mentioned in the Wikipedia page.
neighborhoodList.append('Thavarekere')
neighborhoodList.append('Chikkabanavara')
neighborhoodList.append('Hesaraghatta')
neighborhoodList.append('Jigani')
neighborhoodList.append('Sarjapura')
```

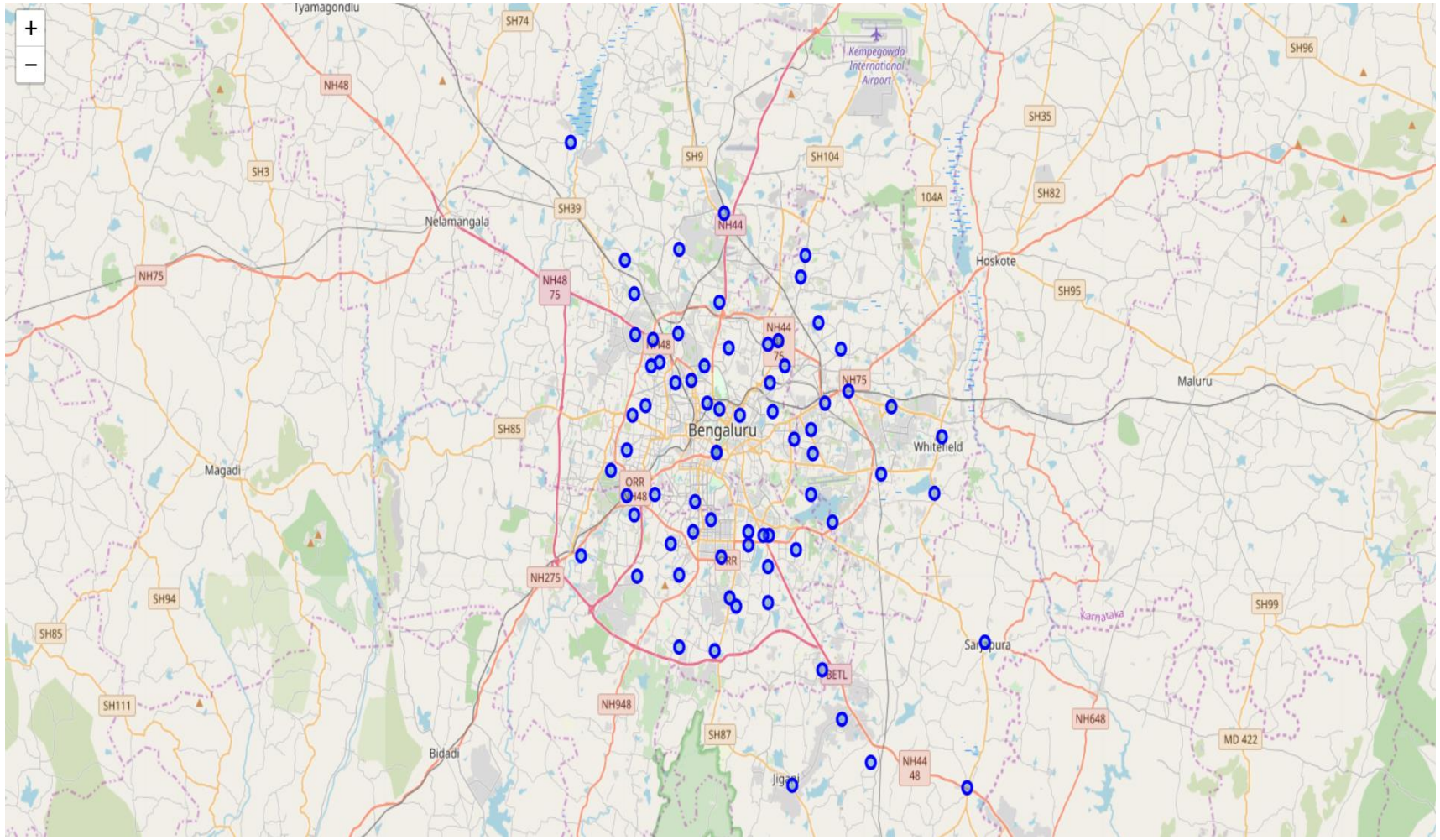
In [267]:
```
# check the neighborhoods and the coordinates
print(blr_df.shape)
blr_df
```

```
(72, 3)
```

Out[267]:

	Neighborhood	Latitude	Longitude
0	Catonment area, Bangalore	12.966180	77.586900
1	Domlur	12.943290	77.656020
2	Indiranagar	12.973940	77.643900
3	Jeevanbheemanagar	12.966010	77.657670
4	Malleswaram	13.006322	77.568416
5	Pete area	12.966180	77.586900
6	Sadashivanagar	13.014830	77.577710
7	Seshadripuram	12.993550	77.579880
8	Shivajinagar	12.987200	77.604010
9	Ulsoor	12.989080	77.627950
10	Vasanth Nagar	12.990730	77.588610

List of all neighborhoods in the city



Display and count of the different kinds of venues that exist in each neighborhood

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Catonment area, Bangalore	12.96618	77.5869	ITC Gardenia	12.967010	77.595618	Hotel
1	Catonment area, Bangalore	12.96618	77.5869	JW Marriott Hotel Bengaluru	12.972362	77.595051	Hotel
2	Catonment area, Bangalore	12.96618	77.5869	UB City	12.971709	77.595905	Shopping Mall
3	Catonment area, Bangalore	12.96618	77.5869	Toscano	12.971980	77.596066	Italian Restaurant
4	Catonment area, Bangalore	12.96618	77.5869	Café Noir	12.971995	77.596001	French Restaurant

Let's check how many venues were returned for each neighorhood

```
venues_df.groupby(["Neighborhood"]).count()
```

Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Anjanapura	8	8	8	8	8	8
Arekere	98	98	98	98	98	98
Attibele	4	4	4	4	4	4
BTM Layout	100	100	100	100	100	100
Banashankari	100	100	100	100	100	100
Banaswadi	86	86	86	86	86	86
Basavanagudi	100	100	100	100	100	100
Basaveshwaranagar	100	100	100	100	100	100
Begur	86	86	86	86	86	86
Bellandur	100	100	100	100	100	100

6. Analyze Each Neighborhood

```
# one hot encoding
blr_onehot = pd.get_dummies(venues_df[['VenueCategory']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
blr_onehot['Neighborhoods'] = venues_df['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [blr_onehot.columns[-1]] + list(blr_onehot.columns[:-1])
blr_onehot = blr_onehot[fixed_columns]

print(blr_onehot.shape)
blr_onehot.head()
```

```
(5418, 192)
```

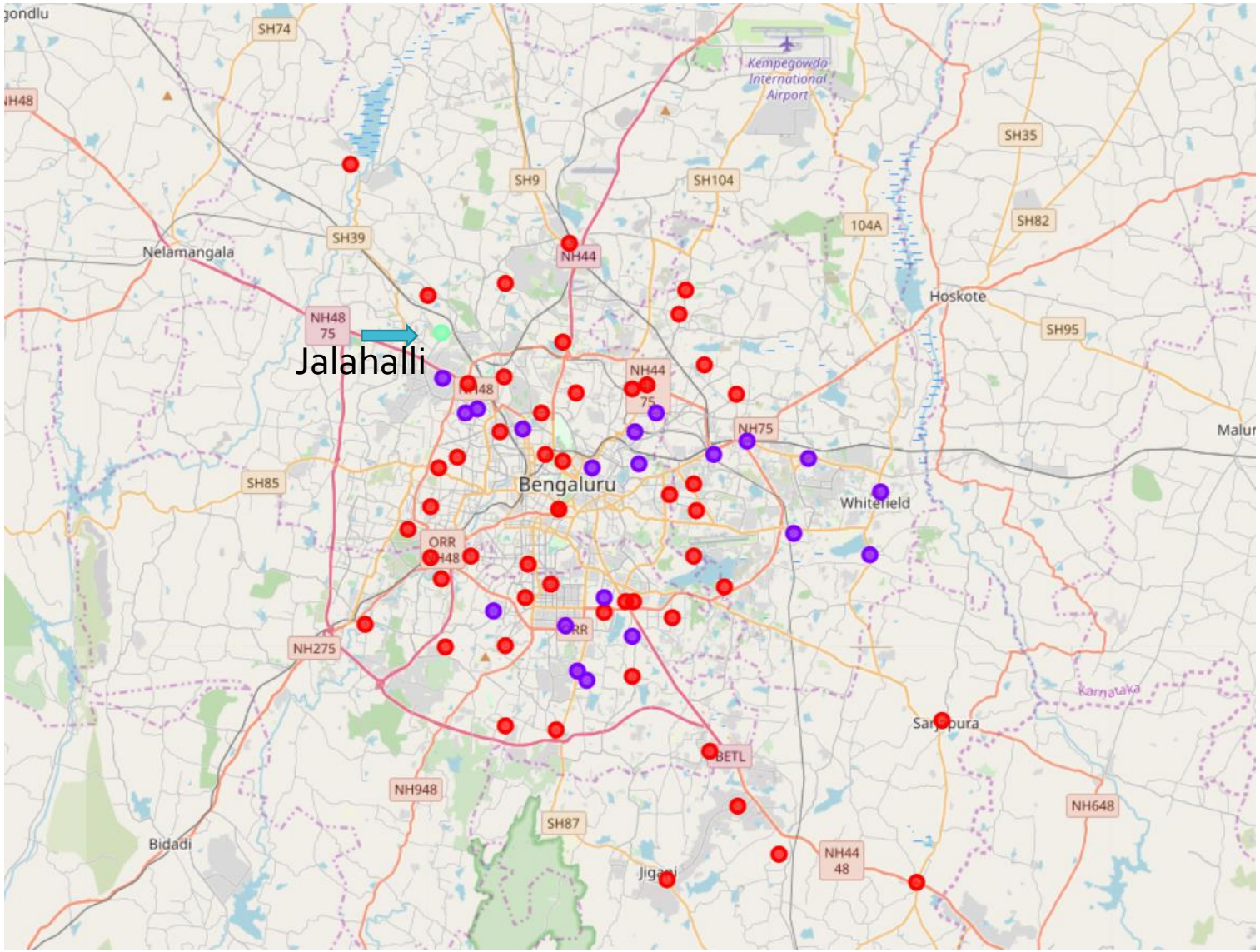
	Neighborhoods	ATM	Afghan Restaurant	Airport	American Restaurant	Andhra Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Auto Workshop	BBQ Joint	B
0	Catonment area, Bangalore	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Catonment area, Bangalore	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Catonment area, Bangalore	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Catonment area, Bangalore	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Unique venues and one hot encoding

```
# print out the list of categories
venues_df['VenueCategory'].unique()[:500]
```

```
array(['Hotel', 'Shopping Mall', 'Italian Restaurant',
       'French Restaurant', 'Japanese Restaurant', 'Lounge',
       'Asian Restaurant', 'Sushi Restaurant', 'Fried Chicken Joint',
       'Deli / Bodega', 'Park', 'South Indian Restaurant', 'Theater',
       'Indian Restaurant', 'Ice Cream Shop', 'Burger Joint', 'Brewery',
       'Cupcake Shop', 'Furniture / Home Store', 'Breakfast Spot', 'Pub',
       'American Restaurant', 'Cricket Ground', 'Seafood Restaurant',
       'Bed & Breakfast', 'Plaza', 'Arcade', 'Gym / Fitness Center',
       'Bookstore', 'Bakery', 'Toy / Game Store', 'Chinese Restaurant',
       'Botanical Garden', 'Snack Place', 'Racetrack', 'Tea Room',
       'Dessert Shop', 'Mexican Restaurant', 'Cocktail Bar', 'Café',
       'Coffee Shop', 'Afghan Restaurant', 'Art Gallery',
       'Parsi Restaurant', 'Sandwich Place', 'Steakhouse', 'Wine Bar',
       'Golf Course', 'Andhra Restaurant', 'Electronics Store',
       'Vietnamese Restaurant', 'Restaurant', 'Soccer Stadium',
       'Hookah Bar', 'BBQ Joint', 'Irish Pub', 'Clothing Store',
       'Mobile Phone Shop', 'Spa', 'Farmers Market',
       'Fast Food Restaurant', 'Chocolate Shop', 'Gaming Cafe',
       'Liquor Store', 'Multicuisine Indian Restaurant', 'Candy Store',
       'Boutique', 'Pizza Place', 'Yoga Studio', 'Gym', 'Trail',
       'Food Truck', 'Lake', 'Karnataka Restaurant', 'Music Venue',
       'North Indian Restaurant', 'History Museum', 'Udupi Restaurant',
       'Department Store', 'Punjabi Restaurant', 'German Restaurant',
       'Butcher', 'Bar', 'Light Rail Station', 'Mediterranean Restaurant',
       "Women's Store", 'Convenience Store', 'Donut Shop',
       'Middle Eastern Restaurant', 'Korean Restaurant', 'Nightclub',
       'Bengali Restaurant', 'Burrito Place', 'Juice Bar', 'Sports Bar',
       'Athletics & Sports', 'Multiplex', 'Bowling Alley',
       'Movie Theater', 'Vegetarian / Vegan Restaurant',
       'Motorcycle Shop', 'Monument / Landmark', 'Gas Station'
```

The clustered set on the map



The clustered set of neighborhoods

Cluster 0

- Anjanapura
- Jigani
- Kalyan Nagar
- Kamakshipalya
- Yelahanka
- Kengeri
- Koramangala
- Kothnur
- Kumaraswamy Layout
- Madiwala
- Bellandur
- Begur
- HSR Layout
- Mathikere
- Nagarbhavi
- Jeevanbheemanagar
- Nayandahalli
- R. T. Nagar
- Rajajinagar
- Rajarajeshwari Nagar
- Ramamurthy Nagar
- Sadashivanagar
- Sarjapura
- Gottigere
- HBR Layout
- Girinagar
- Seshadripuram
- Uttarahalli
- Vasanth Nagar
- Vidyaranyapura
- Vijayanagar
- Pete area
- Jayanagar
- Yeshwanthpur
- Domlur
- Basaveshwaranagar
- Electronic City
- Hebbal
- BTM Layout
- Bommasandra
- CV Raman Nagar
- Basavanagudi
- Banaswadi
- Catonment area, Bangalore
- Indiranagar
- Chandrapura
- Chikkabanavara
- Banashankari
- Hesaraghatta
- Attibele
- Horamavu

o

Cluster 1

- **Whitefield**
- **Arekere**
- **Varthur**
- **Thavarekere**
- **Ulsoor**
- Malleswaram
- Mahalakshmi Layout
- Mahadevapura
- Shivajinagar
- J. P. Nagar
- Nandini Layout
- Peenya
- Padmanabhanagar
- Marathahalli
- Lingarajapuram
- **Krishnarajapuram**
- **Hoodi**
- **Hulimavu**
- **Bommanahalli**
- **Kammanahalli**

Cluster 2

- **Jalahalli**

Conclusion and future directions

- Used k-Means clustering to predict the best neighborhoods to construct new Shopping malls for profit maximization of property developers
- Accuracy of the model has room for improvement.
- More features can be included for enhanced predictions
- Ideas include:
    - 1. Income of people.
    - 2. Diaspora of neighborhood.
    - 3. A different and more accurate clustering algorithm.